Parallel Graph Algorithms for Complex Networks

Edmond Chow,
Tina Eliassi-Rad, Keith Henderson

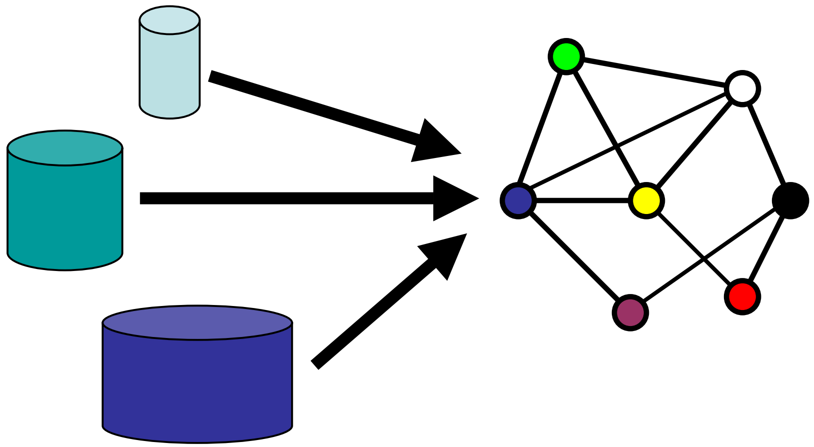
Lawrence Livermore National Laboratory
Comp/NAI LDRD Project

UCRL-PRES-203191

DHS Advanced Scientific Computing PI Meeting
March 16, 2004

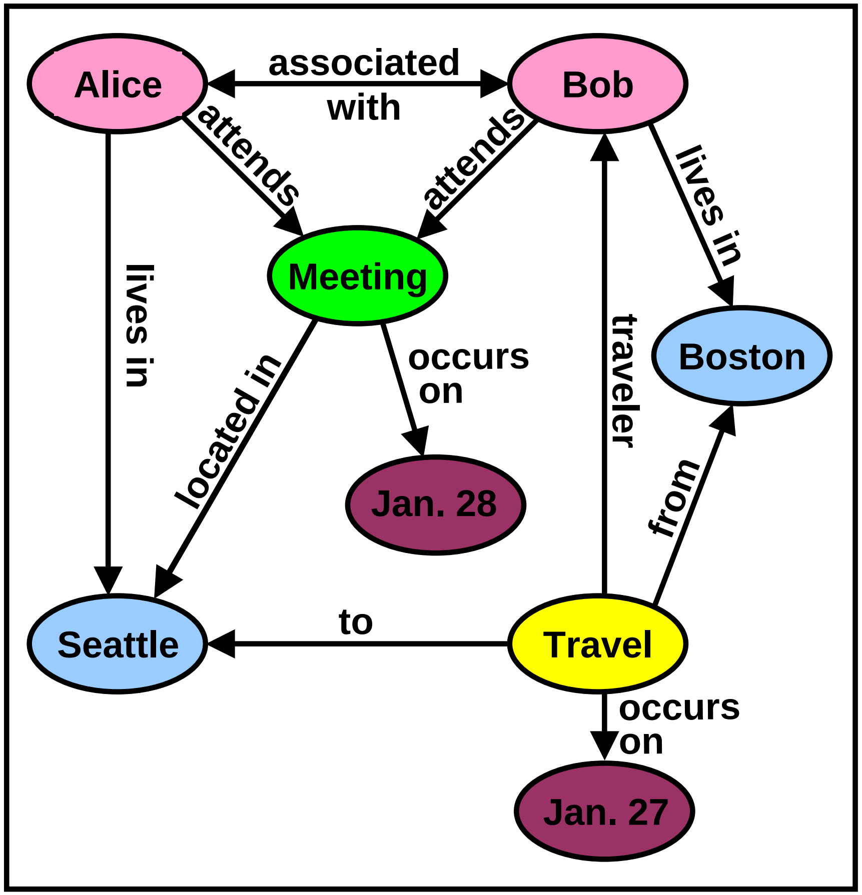
Intelligence analysts must find relationships in huge amounts of data

- **Data is collected from multiple sources at increasing rates**
- **Challenge: identify relationships and uncover patterns in a timely manner**
- **Approach: use *semantic graphs* to represent the data and graph algorithms to discover hidden relationships**

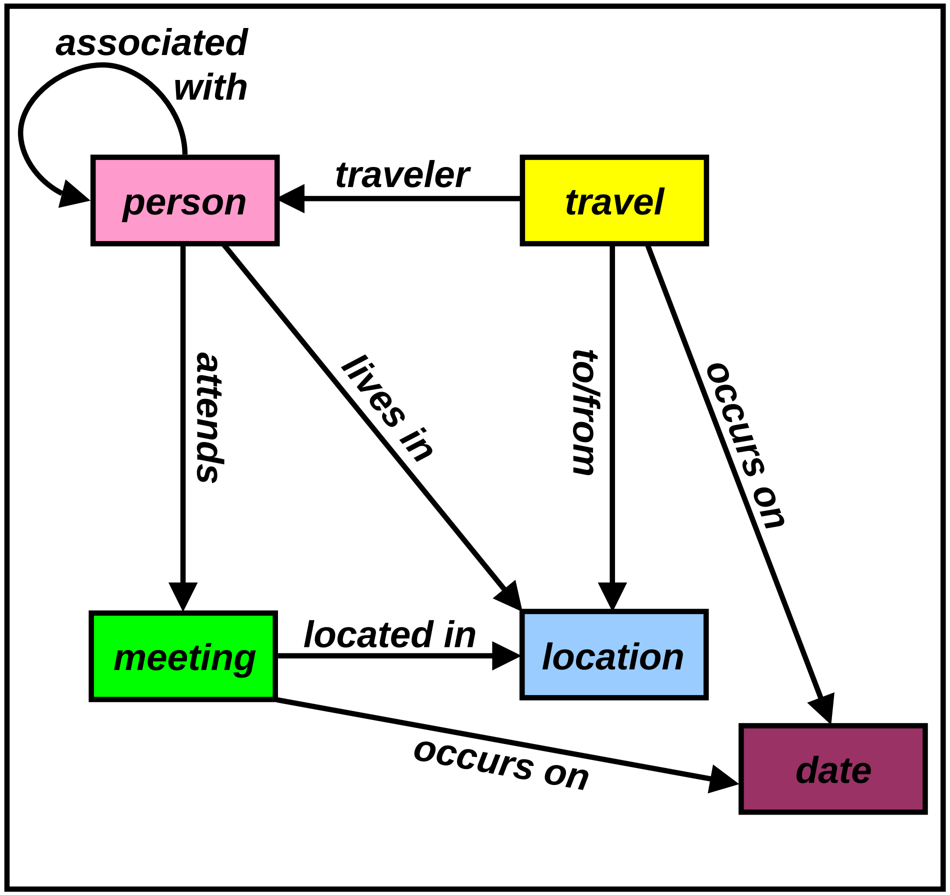


Semantic graphs have types and attributes on the vertices and edges

Semantic Graph



Ontology

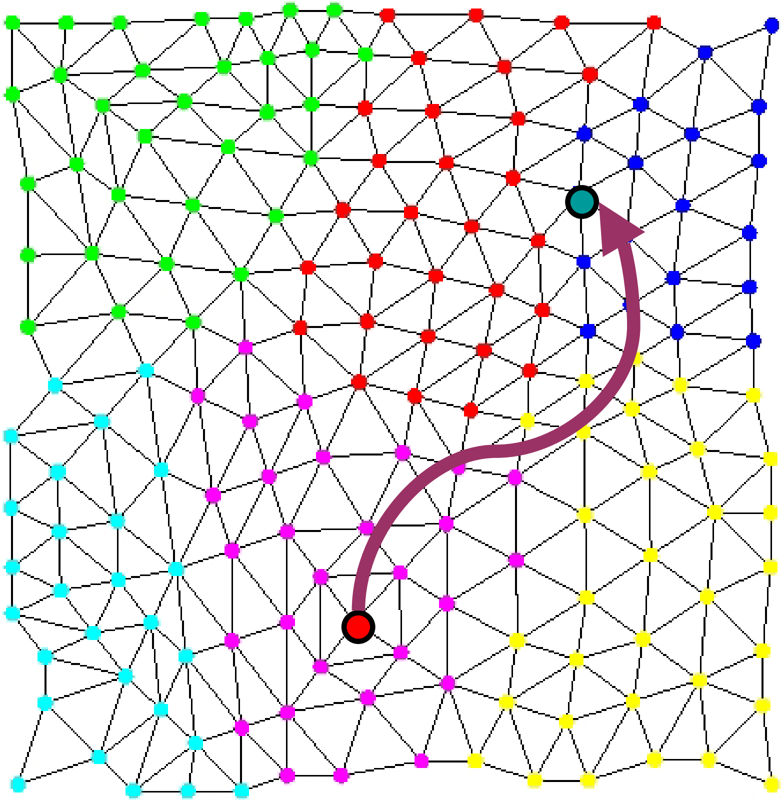


Semantic graphs in intelligence applications are becoming enormous

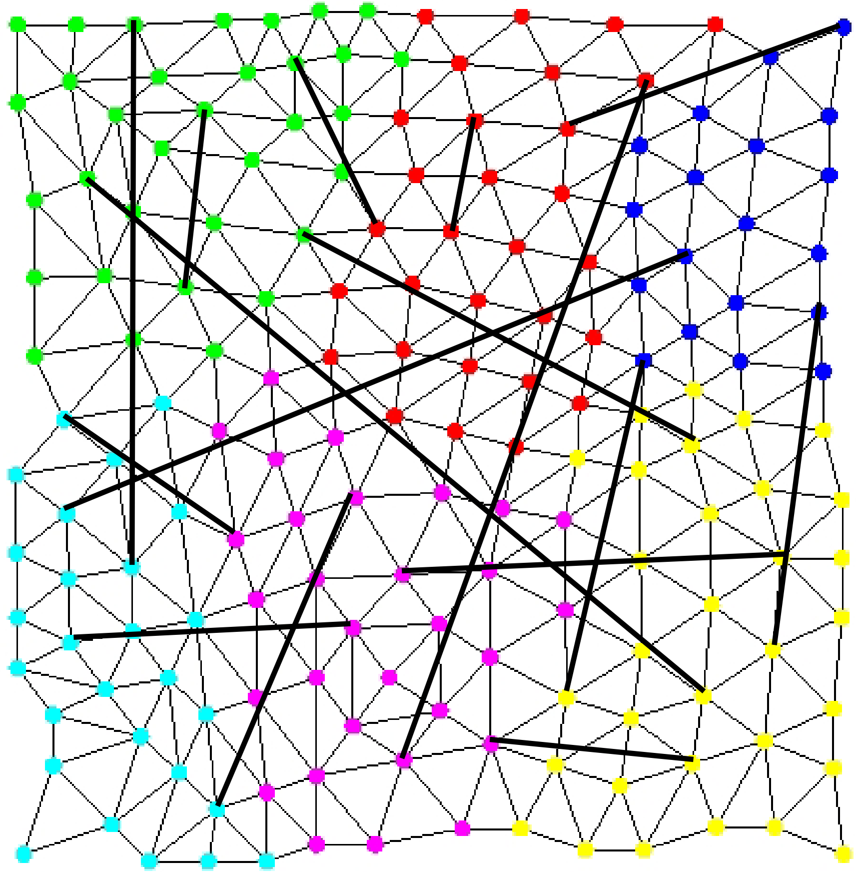
- Distributed memory parallel computers must be used to store and search these graphs
- Graphs must be partitioned onto separate memories and graph searches must have low communication cost
- Semantic graphs have topological properties that make partitioning very challenging
- Our goals are to develop partitioners and efficient parallel search algorithms

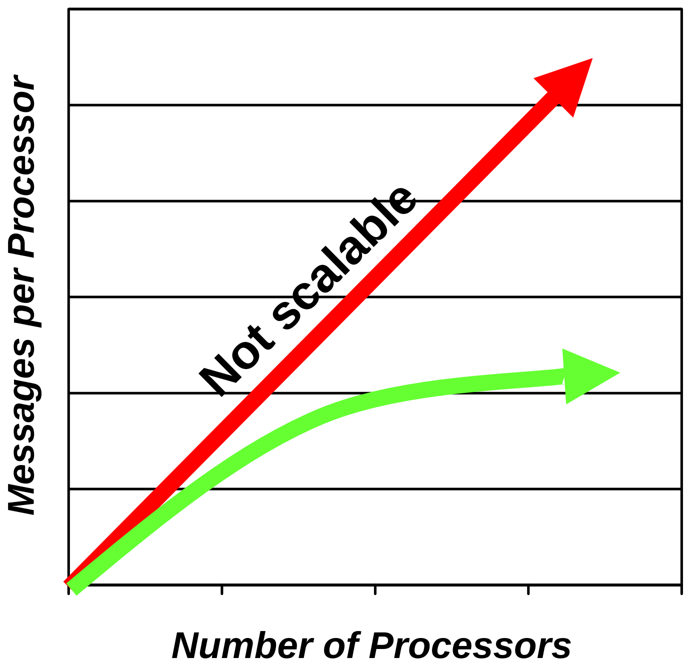
We are developing parallel algorithms to search massive graphs

- Semantic graph models
- Partitioning approaches with bounded messaging cost that utilize the ontology
- New ontology-based probabilistic heuristics for searching graphs
- New heuristics for searching general graphs

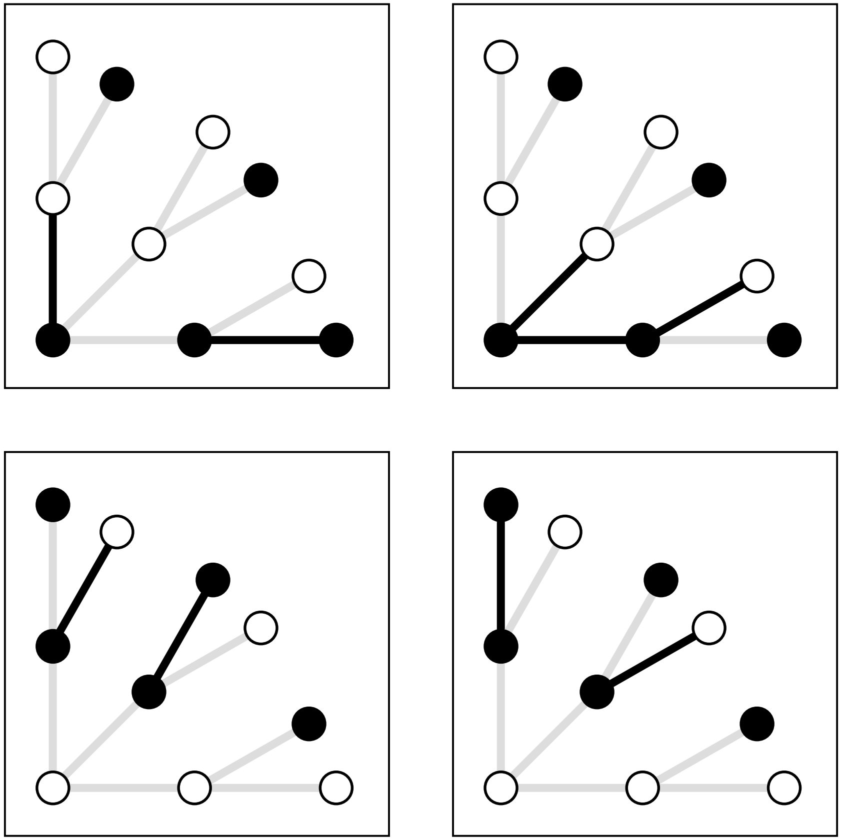


Partitions of a complex network may induce $O(P^2)$ communication





We use an edge partitioning approach to get $O(P)$ communication



Example partitioning on 4 processors

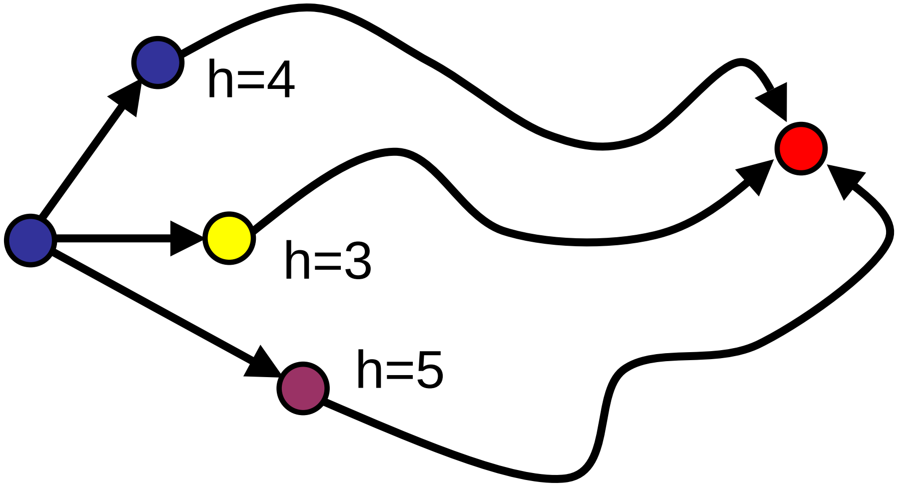
- Coarse partitioning of $\sqrt{P}$ parts
- Expanding a vertex is $\sqrt{P}$ communication
- Updating the fringe is also $\sqrt{P}$ communication
- Disadvantage is more complicated parallel implementation
- Edge partitioning can be based on edge types

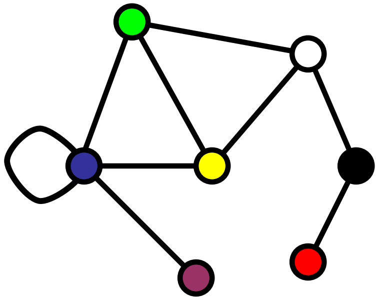
2D-hypergraph partitioning shows dramatic potential of this approach

	Vertices	Edges	Num. Messages 1D	Num. Messages 2D	Comm. Volume 1D	Comm. Volume 2D
Mesh	205761	615360	2000	160	2136	2296
Random	205761	615360	157200	760	158668	159576
Scale-Free	205761	617277	156800	760	201786	221174
Spatial	51682	154080	22400	760	23126	23922
Web	325729	1090108	20800	680	31727	44246

Simulated communication performance for 400 partitions shows potential for orders of magnitude improvement with this approach

We are developing heuristics to guide and accelerate the search





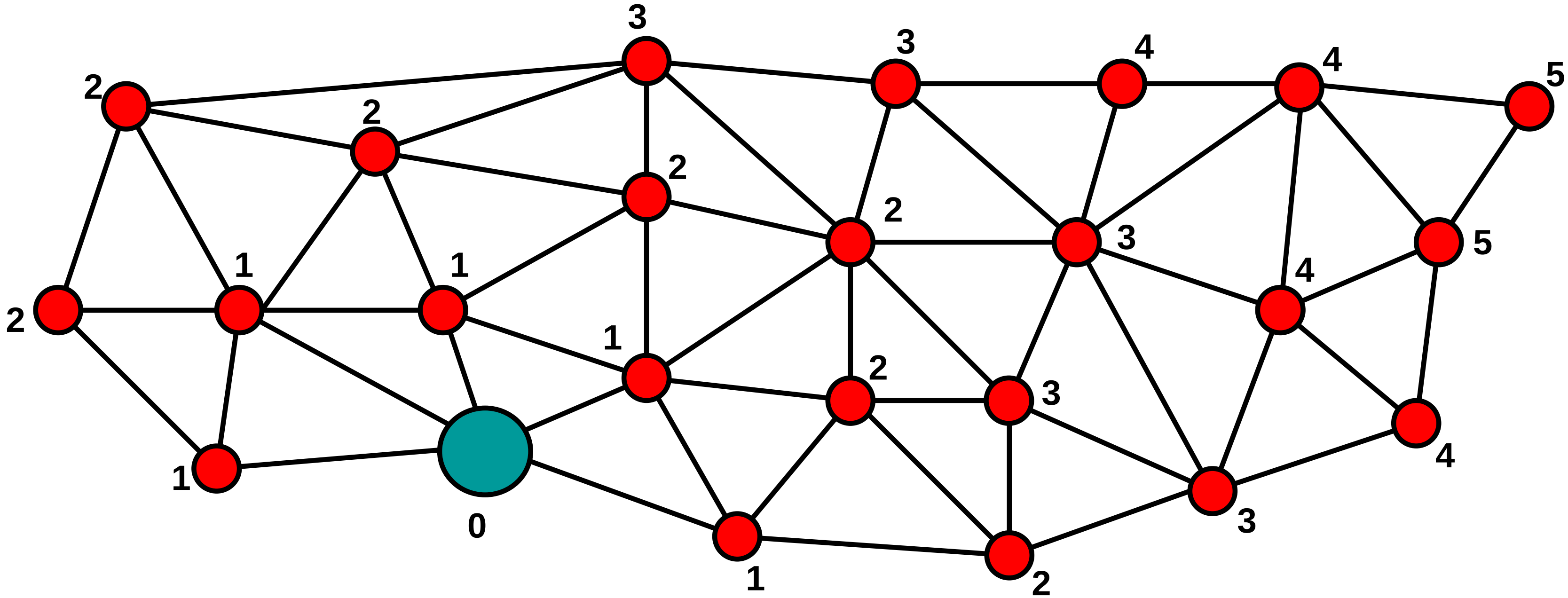
The ontology gives a lower bound on distance

- **Distance estimates are used to guide the heuristic search (A* search)**
- **Vertex and edge frequencies give a probabilistic measure of the heuristic estimate**
- **Parallelization will use standard owner-computes model**

For typical ontologies, heuristic search is 2-3 times faster than BFS

Ontology edge density	***BFS vertices visited***	***Ontology heuristic work ratio***	***Probab. heuristic work ratio***
0.05	26594	0.483	0.323
0.25	26443	0.655	0.495
0.50	24533	0.798	0.634
0.75	22534	0.885	0.676
1.00	23924	1.000	0.779

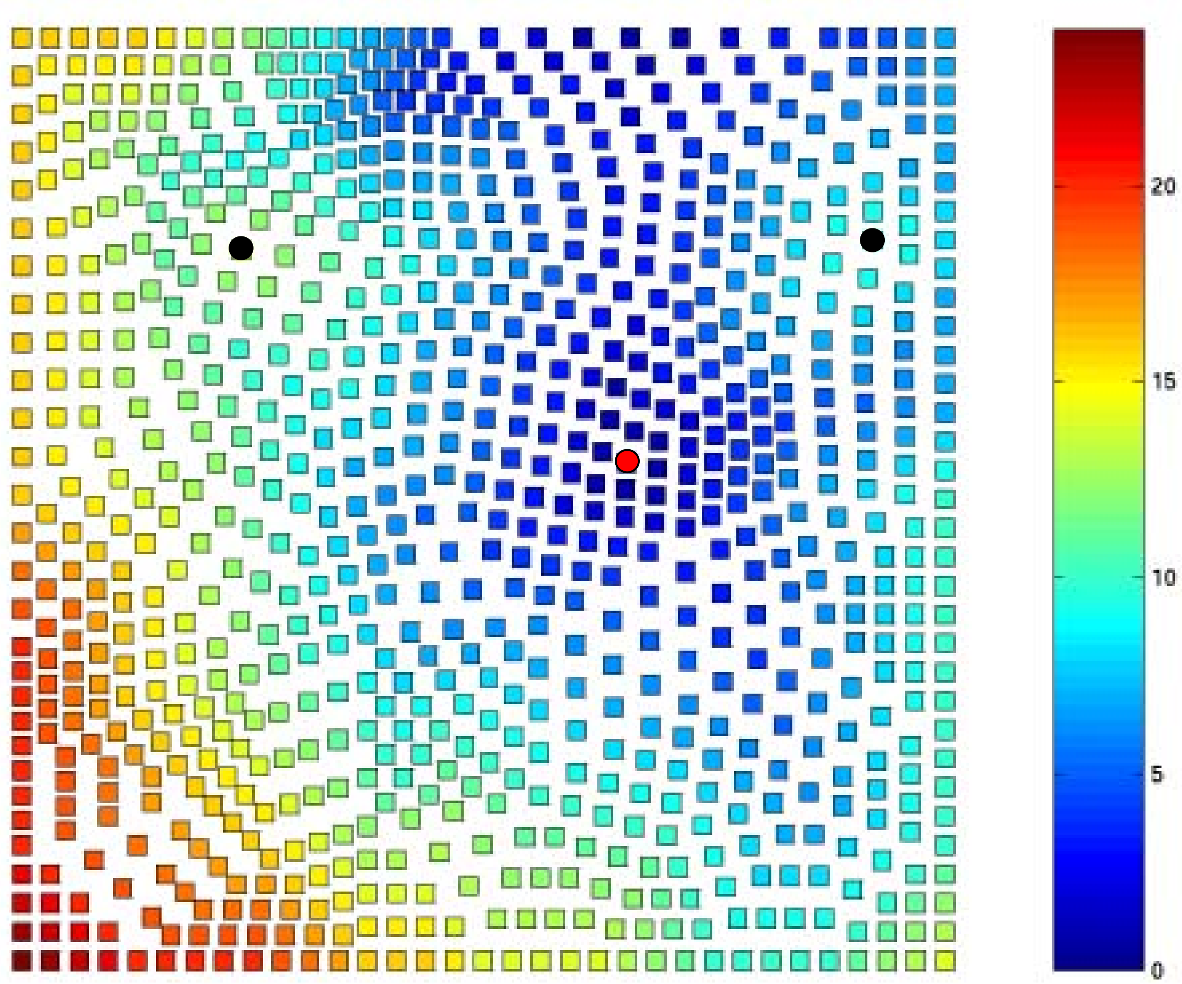
We are also developing heuristics based on the graph itself



$h_1(i,j) = |\ \mathrm{level}(i) - \mathrm{level}(j)\ |$

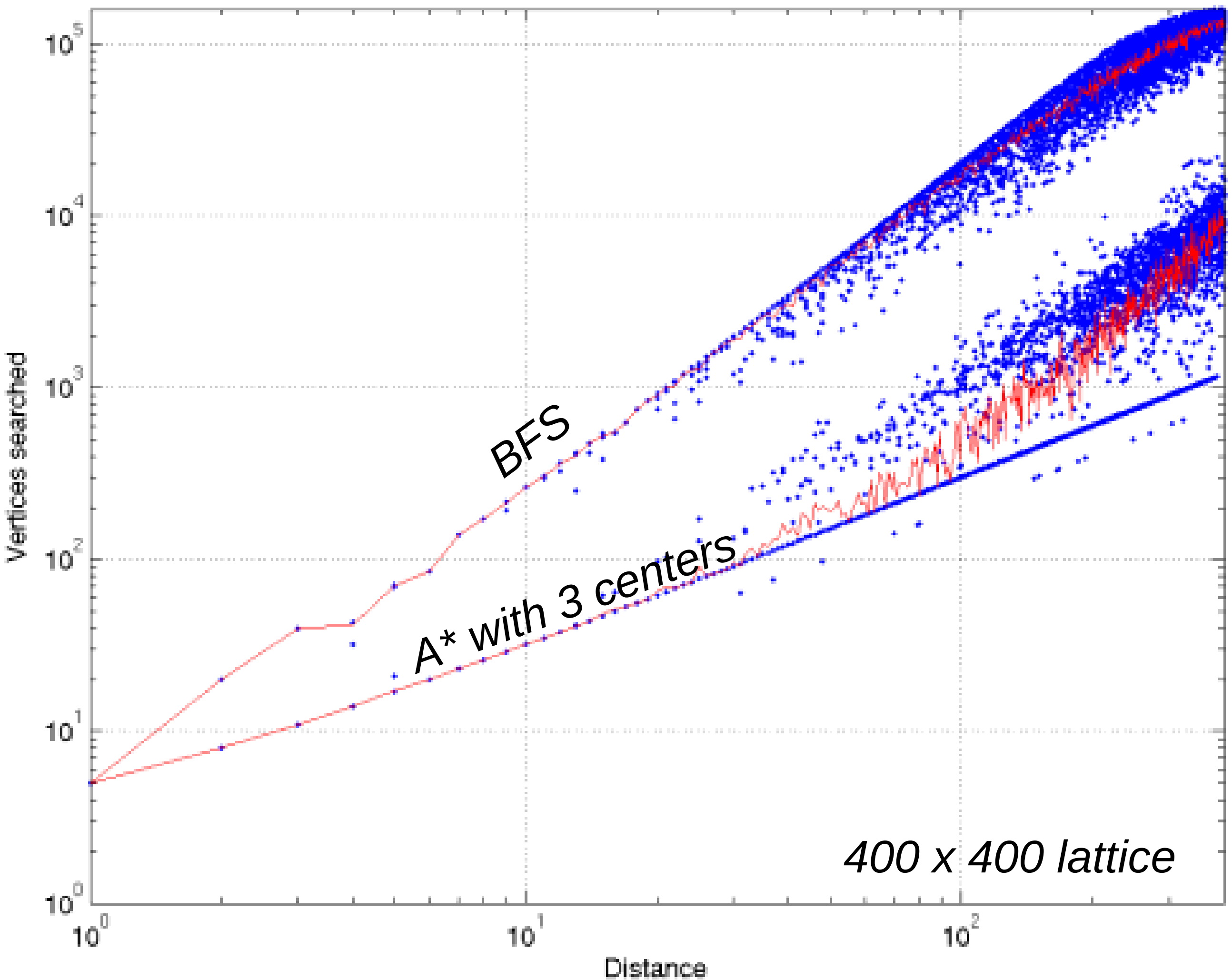
$h(i,j) = \max\{\ h_1(i,j), \ldots, h_m(i,j)\ \}$, m centers

Example: heuristic distance to a vertex on a mesh

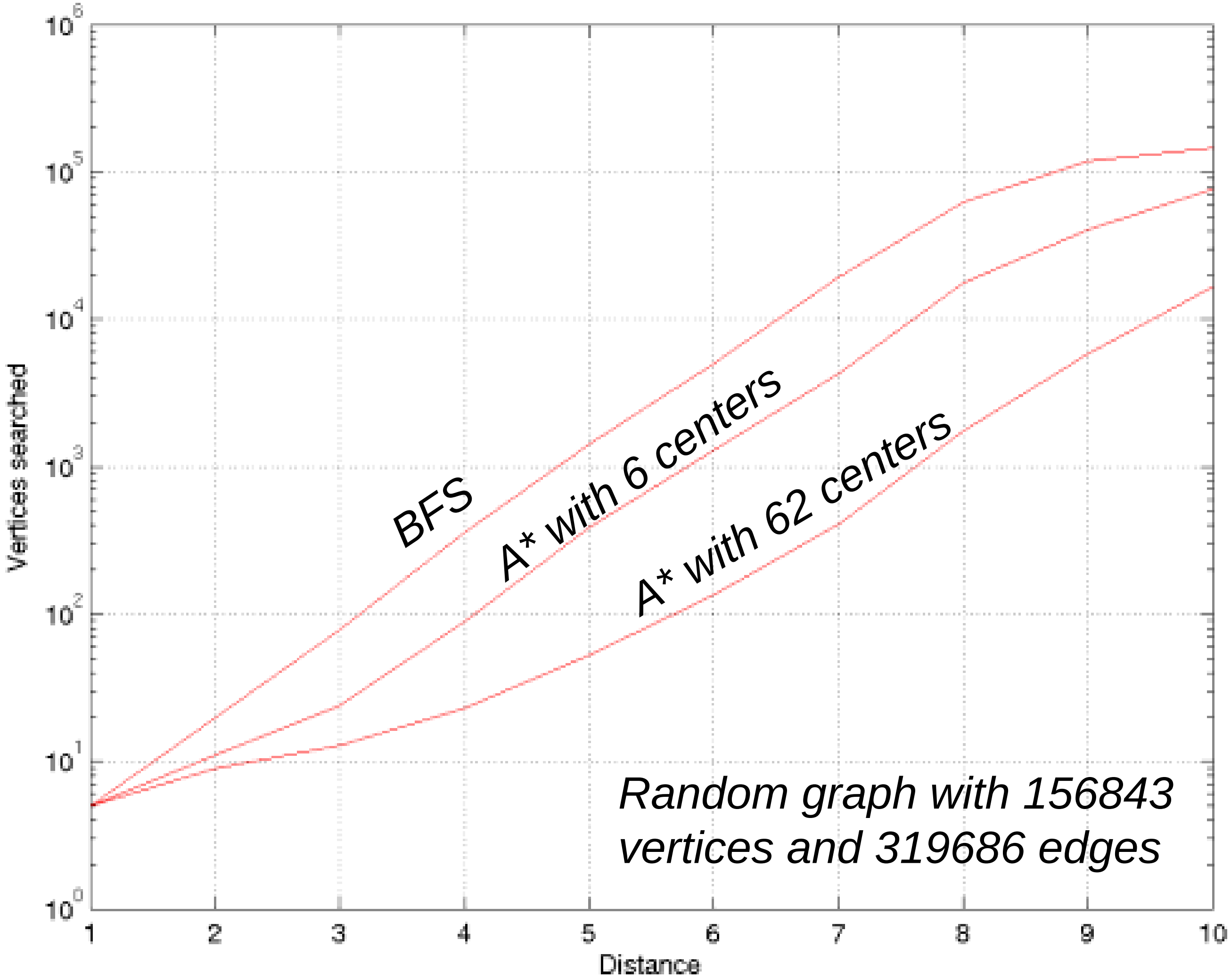


Heuristic distance to red vertex given two center vertices

For a lattice graph, search complexity reduces from $O(L^2)$ to $O(L)$



For a random graph, search time can be reduced ten-fold



Our new algorithms will give graph analysis vastly more power

- **Enable search on distributed parallel computers**
- **Anticipate significant decreases in communication time and search complexity**

- **Future research will continue to be relevant to large-scale applications**
  - **Develop parallel implementations**
  - **Exploit properties of semantic graphs and complex networks in 2D partitioning**
  - **Develop hierarchical graph representations and algorithms**
  - **Develop algorithms for dynamically changing graphs**
  - **Exploit temporal locality in queries to drive partitioning**

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.